

Title: Relationship between Loan Interest Rate and Properties of the Loan and Borrower
By Jonathan Fivelsdal

Introduction:

In the wake of the U.S. economic recession of 2007-2009 and the subprime mortgage crisis, financial institutions adopted stricter credit standards which made it more difficult for individuals and businesses to receive credit and loans. Peer-to-peer lending allows individuals to lend money directly to other individuals. Over the past few years, peer-to-peer lending has gained popularity as an alternative means of obtaining loans. In a report on the availability of credit to small businesses, it states that data provided by Propser.com and Lending Club.com indicates that the dollar volume of peer-to-peer lending grew by nearly 300 percent between 2008 and 2011 and that the total loan dollars that go to small business has been increasing steadily since 2009 [1]. In this analysis we examine relationships between the interest rate and other characteristics of peer-to-peer loans.

Methods:

Data Collection

The data used for this analysis consist of a sample of 2,500 peer-to-peer loans issued through the Lending Club. The data were downloaded on November 8th 2013 using the R programming language.

Statistical Methods

We used a t-test to compare the average interest rate at different fico score levels [2]. We also used a multiple linear regression equation in order to explore relationships between the interest rate and other variables within the data set [3].

Results:

According to the credit agency Experian, a credit score of above 700 usually suggests good credit management [4]. In our analysis, we split up the borrowers that have a high credit rating which we define as a credit rating of 700 or more from those in the low credit score category which is having a credit score below 700. Borrowers that have a fico range where the lower endpoint of the fico range is 700 or higher are considered to be in the high credit score group and those that have a fico range where the lower endpoint is less than 700 are considered to be in the low credit score group.

We performed a two sample, unpaired t-test involving the loan interest rates associated with 1,292 borrowers with a high credit rating versus the 1,208 borrowers with a low credit rating. The null hypothesis for the t-test is that the average interest rate is the same for both the low credit score group and the high credit score group. The alternative hypothesis is that the mean interest rate for the low credit score group is not equal to the mean interest rate of the high credit score group.

The results of the t-test were a t value of 40.5529, degrees of freedom of 2,492.862, and a p-value approximately equal to zero. The absolute value of our t-value is 40.5529 and is greater than the 0.05 significance critical value of 1.9673 and so at the 0.05 significance level we reject the null hypothesis. The sample estimate of the mean interest rate for consumers with a low credit rating is 15.79 % while the estimate of the mean interest rate for consumers with a high credit rating is 10.52%. The mean interest rate for both credit score groups is plotted in Figure 1. The 95% confidence interval for the difference in mean interest rates between consumers with low credit scores and those that have high credit scores (95% CI: 5.01% - 5.52%). Our low p-value ($P < 0.0001$) indicates a statistically significant difference in the mean interest rate between consumers that have a low credit rating versus those that have a higher credit rating.

We fit a linear regression model relating interest rate to several variables in the data set. A simple linear regression model only explained a small amount of the variation in the data. The following model was used in order to improve the percentage of variation explained in the model and also to address potential confounders.

$$Y = 11.3159 - 3.4711 X1 + 0.2524 X2 + 0.0284 X3 + 3.9498 X4 + 0.0447 X5 + 3.0987 X6 + 0.1055 X7 - 1.1753 X8 + 0.0036 X9 + 0.3328 X10 + 2.4218 X11 + 0.1366 X12 + e$$

Where Y is the interest rate, X1 is the fico score category, X2 is the number of credit inquiries in the last six months, X3 is the debt to income ratio, X4 is the employment length, X5 is the homeownership term, X6 is the loan length, X7 is the loan purpose, X8 is the open credit lines term, X9 is the monthly income, X10 is the revolving credit term, X11 is the amount requested term, X12 is the amount funded by investors, and e is the error term. The inquiries in the last six months (X2) and the debt to income ratio (X3) are continuous variables while the other 10 predictor variables are categorical variables.

The fico score category (X1) is a binary factor, where 1 represents a credit score above 700 and 0 represents a credit score below 700. The employment length (X4) is also a binary factor, where 1 represents being employed for ten years or more and 0 represents being employed for less than ten years. The homeownership term (X5) is a binary factor such that 1 represents a person that owns their home and 0 represents someone with a mortgage or is renting. The loan length term (X6) is a binary factor such that 1 represents a loan term of more than 36 months (3 years) and 0 represents a loan length of less than 3 years. The loan purpose term (X7) is a binary factor, such that 1 represents a small business loan and 0 represents any other type of loan. The open credit lines variable (X8) is a binary factor such that 1 represents having five or more open credit lines and 0 represents having less than five open credit lines. The monthly income term (X9) is a binary factor such that 1 represents a monthly income of \$3000 or more, and 0 represents a monthly income of less than \$3,000. The revolving credit balance term (X10) is a binary factor such that 1 represents having a revolving credit balance of more than \$10,000 and 0 represents having a credit balance of less than \$10,000. The amount requested term (X11) is a binary factor such that 1 represents an amount requested that is greater than or equal to \$30,000 and 0 represents an amount requested that is below \$30,000. The amount funded by investors term (X12) is a binary factor such that 1 represents a funded amount of \$30,000 or more and 1 represents a funded amount of less than \$30,000.

All variables except for homeownership (X5), loan purpose (X7), monthly income (X9), and the amount funded by investors (X12) have a low p-value ($P < 0.0001$) and are statistically significant at a 0.05 significance level. The p-values for homeownership, loan purpose, monthly income, and amount funded by an investor are 0.78, 0.65, 0.98, and 0.80 respectively. The model has a multiple R-squared value of 0.74 which indicates that the model explains 74% of the variation within the data. Within the dataset there is 1 missing value for monthly income and 2 missing values for the open credit lines, revolving credit balance, and inquiries in the last six months respectively. Missing values were excluded from regression calculations.

We also considered the relationship between interest rate and geographic region. The states were divided up into regions according to how the United States Census Bureau defines the Northeast, Midwest, South, and West (information on how these regions are defined can be found here [5]). The mean interest rate was calculated among the states to be 12.97% for the Northeast, 13.05% for the Midwest, 13.12% for the South, and 13.10% for the West. We also calculated the average of the lower endpoint of the fico range across states. The average of the fico scores were 706.59 for the Northeast, 706.02 for the Midwest, 706.51 for the South, and 704.52 for the West.

Conclusions: Our analysis suggests that there is a significant relationship between the interest rate of a peer-to peer loan and variables such as a borrower's FICO score, employment length, the length of the loan, the number of open credit lines, the revolving credit balance of the borrower and the amount requested by the borrower. Our analysis using the regression model also indicates that factors such as home ownership, loan purpose, monthly income, and the amount funded by an investor did not have a statistically significant relationship to the interest rate of the loan. In addition, there did not seem to be much of a relationship between the interest rate of a loan and geographical location. The mean interest rates and mean FICO scores remained fairly constant across geographical regions. The model adjusts for several factors although potential confounders such as the age and gender of the borrower may not be adjusted for in the regression model. Further study may include adjusting for other factors not in the model.

References:

- 1. Board of Governors of the Federal Reserve System. Report to Congress - The Availability of Credit to Small Businesses, September 2012.**
URL:
<http://www.federalreserve.gov/publications/other-reports/availability-of-credit/September-2012-Providers-of-Credit-to-Small-Businesses.htm>.
Accessed 11/15/2013
- 2. NIST/SEMATECH e-Handbook of Statistical Methods. URL:**
<http://itl.nist.gov/div898/handbook/eda/section3/eda353.htm>. **Accessed 11/15/2013**
- 3. Field A, Miles J and Field Z (2013). Discovering Statistics Using R. SAGE Publications Inc., Thousand Oaks, California. ISBN 978-1-4462-0046-9.**
- 4. Experian. What Is a Good Credit Score? URL:**
<http://www.experian.com/credit-education/what-is-a-good-credit-score.html>. **Accessed 11/8/2013**
- 5. 2007 Economic Census - Regions and Division. URL:**
http://www.census.gov/econ/census07/www/geography/regions_and_divisions.html.
Accessed 11/16/2013

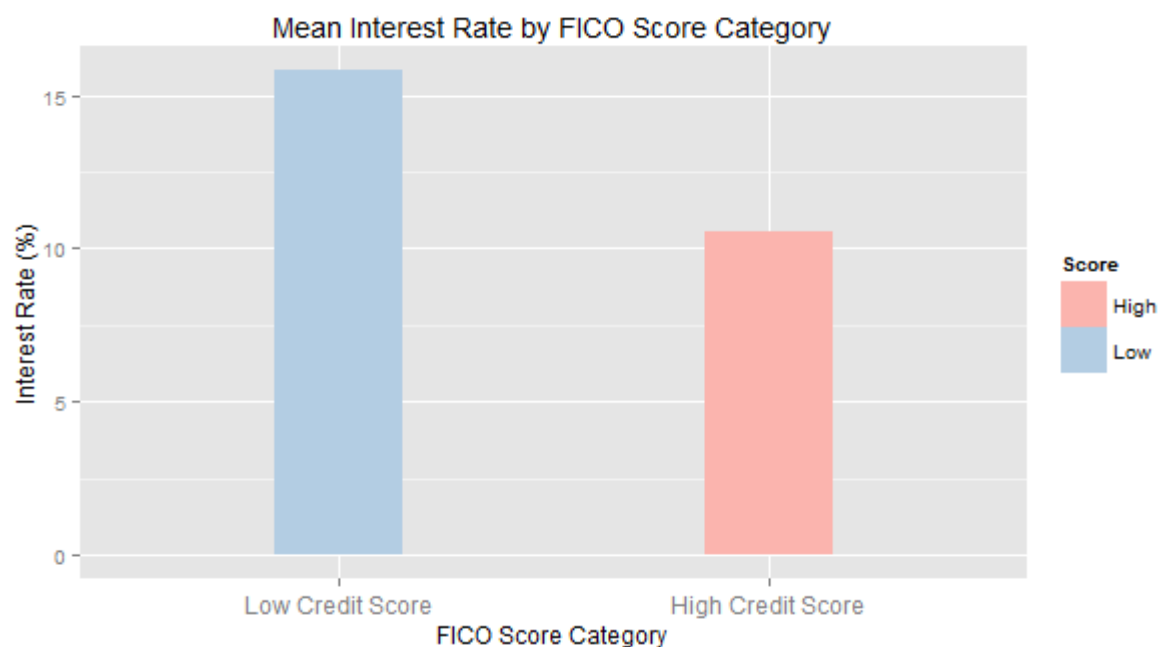


Figure 1 A bar graph of the mean interest rate for borrowers with a low credit score and the mean interest rate for borrowers with a high credit score. The high credit score category is defined as containing FICO score ranges such that the lower endpoint of the FICO score range is 700 or higher and the low credit score category is defined as containing FICO score ranges such that the lower endpoint of the FICO score range is below 700. The average loan interest rate is a little more than 5% higher for the group of borrowers that have a low credit score versus those borrowers with a higher credit score. A t-test was performed and the resulting test statistic was compared to the critical value at a 0.05 significance level and suggests that the mean interest rates of the two groups are unequal.